EP3260: Machine Learning Over Networks
Computer Assignment 7
Due Date: Apr 15, 2023

Computer Assignment 7 - Deep Neural Networks

Consider the "MNIST" dataset, and a DNN with J layers and $\{N_j\}$ neurons on layer j.

a) Train DNN using SGD and your choices of hyper-parameters, L, and $\{N_j\}_{j\in[J]}$. Report the convergence rate on the training as well as the generalization performance. Feel free to change SGD to any other solver of your choice (give explanation for the choice).

b) Repeat part a) with mini-batch GD of your choice of the mini-batch size, retrain DNN, and show the performance measures. Compare the training performance (speed, accuracy) using various adaptive learning rates (constant, diminishing, AdaGrad, RMSProp).

c) Consider design of part a) and fix $\sum_j N_j$. Investigate shallower networks (smaller J) each having potentially more neurons versus deeper network each having fewer neurons per layer, and discuss pros and cons of these two DNN architectures.

d) Split the dataset to 6 random disjoint subsets, each for one worker, and repeat part a) on master-worker computational graph.

e) To promote sparse solutions, you may use l_1 regularization or a so-called dropout technique. Explain how you incorporate each of these approaches in the training? Compare their training performance and the size of the final trained models.

f) Improving the smoothness of an optimization landscape may substantially improve the convergence properties of first-order iterative algorithms. Batch-normalization is a relatively simple technique to smoothen the landscape [1]. Using the materials of the course, propose an alternative approach to improve the smoothness. Provide numerical justification for the proposed approach.

References

[1] S. Santurkar, D. Tsipras, A. Ilyas, and A. Mądry, "How Does Batch Normalization Help Optimization?" in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, ser. NIPS'18. Red Hook, NY, USA: Curran Associates Inc., 2018, p. 2488–2498.